Response to Referee #2

We would like to thank the three reviewers for their helpful comments and corrections. We propose a revised version of the paper in which we have taken into account their comments and suggestions. We hope that our revisions will address their expectations.

In the following document, the reviewer's comments are shown in **black** and our point-by-point responses are in **blue**.

The main adjustments proposed in the new version of the manuscript are described below:

- The section "Code and data availability" has been modified to add a brief description of the ICOLMDZ model.
- We have reorganized the figures and tables to improve the flow as reviewers suggested. The figure and table numbers have changed from the initial version of the manuscript:
    - Figure 5 has been removed to avoid repetition with Table 5.
    - Table 3, Table 4, Figure 3 and Figure 8 have been moved to the Supplementary Material.
    - Table 6 has been removed.
    - Figures S2 and S4 were moved from the Supplementary Material to the main text.
    - For each tendency, the results obtained during the first learning and those obtained with the second learning were separated to ease the flow of the presentation of the results. The results of the second learning are only shown when they are discussed (in Section 5 "Refined approach: integrating physical knowledge"). The revised figure numbering, for the zonal wind tendency, is as follows: Figure 4 has been split into Figure 3 and Figure 9, and Figure 6 becomes Figure 4 and Figure 10. We did the same for the five other tendencies in the Supplementary Material. We also separated the results from Table 5 into two tables.
- We propose a new colour scheme for vertical profiles according to the Coblis-Color Blindness:

	Initial training	**Second training with laplacians**
DNN	Green with solid lines	Blue with solid lines
U-Net	Orange with dashed lines	Red with dashed lines

- Section 6 "Discussion" has been merged with Section 7 "Conclusions". The results of this merger is Section 7 "Conclusions and discussion".

- As suggested by the reviewer 3, new results on the emulation of a realistic configuration have been added to the manuscript in Section 6 “Results for a realistic setup”. The data used for this realistic configuration are described in Section 2.1 “ICOLMDZ simulation data”. The terms “for the aquaplanet setup” have been added to the names of Sections 4 “Initial training performance for the aquaplanet setup” and 5 “Refined approach for the aquaplanet setup: integrating physical knowledge” to avoid confusion.

The manuscript presents the development of data-driven parameterizations for the LMDZ Atmospheric General Circulation Model, specifically in an aquaplanet configuration.

In the first experiment, the authors utilize low-level model variables (e.g., temperature, humidity) to emulate the entire subgrid-scale physics. They train two neural networks (NNs): a simple feedforward neural network (DNN) and a UNet with residual blocks. Both NNs struggle to capture sufficient variance in the output variables. The authors attribute this limitation to an inadequate representation of turbulence, particularly in the boundary layer. To mitigate this issue, they incorporate laplacians of part of the variables among the NN input variables, which may enhance the NNs' ability to represent turbulence. This modification leads to improvements in the R^2 score and the variance of the NN outputs.

The study yields two primary findings:

1. The UNet architecture performs better than the feedforward NN.
2. Incorporating additional variables, thereby embedding physical knowledge into the NNs, can improve their performance.

Overall, the study is comprehensive and well-documented. However, before recommending its publication, I have several suggestions for refinement.

General comments.

- There are many interesting results in this paper. However, sometimes I was lost in the details and between figures and tables. There are repetitions and reminders of previous results, which make the paper hard to follow. For example, Figure 5 and Table 5 ultimately provide the same information. Tables 3 and 4, as well as Figure 3, could possibly be moved to the Supplementary Material. Additionally, it might be useful to merge Table 4 and Figure 3 by including the total number of parameters on Figure 3. Combining Tables 1, 2, and 6 into a single table would also streamline the presentation. These are suggestions, and the authors can take them or not.

  We thank the reviewer for this advice. As suggested, we have:

- removed Figure 5,
- moved Table 3, Table 4 and Figure 3 in the Supplementary Material,
- added the total number of parameters on Figure 3,
- removed Table 6.

- Please proofread. The grammar can be odd in places.

  We have now proof-read the manuscript better.

- Please revise the scale of the colormaps used. The colors are too saturated in Figure 6 (and some of the Figures in the Supplementary) making it difficult to highlight the results.

  In the revised version of the paper, we have applied a logarithmic scale to improve the readability of the results for the cross-sections.

- Your extrapolation regarding execution times might be correct, but in my experience, estimating the gain when using the NN instead of the physical model is much more complex. It also depends strongly on the strategy chosen to implement the NN. I would not risk providing too many details on this subject.

  We totally agree with this comment. Accordingly, we have modified the paragraph concerning the execution time estimates by adding the following sentences: "However, we would like to emphasize that estimating the gain when using an emulator instead of a physical model is much more complex. This is why these estimates should be considered as a preliminary test. It remains to be seen whether this gain would remain when the emulators are coupled with the dynamical core of the model, i.e., in an online setup. For instance, there is generally a substantial overhead when exchanging data from the dynamical core run on CPUs to the physics emulators run on GPUs. One way to alleviate this issue would also be to run the dynamical core on GPUs, which is already possible with the latest development of DYNAMICO".

Point comments.

- Eq. (2): Is this an approximation?

  Yes, it is an approximation. To clarify, we have decided to remove this equation and the corresponding sentences, on lines 134-135.

- line 231: Is it a 1D UNet? I recommend revising the first sentence of the paragraph to state that you used (rather than developed) the UNet architecture.

Yes, it is a 1D U-Net, we are using 1D convolutions. This information is now clearly added to the first sentence of the paragraph. We have also replaced "developed" by "used" as advised in this sentence. Here is the revised sentence: "We have also used a specific architecture, called a U-Net model (Ronneberger et al., 2015), where we use 1D convolutional layers".

- line 315: Even though I ultimately agree with you on choosing the UNet instead of the DNN, I find that this is not immediately 'obvious'.

  We have removed the terms "it is obvious that".

- line 380: 'It appears that the U-Net architecture has a better ability to capture the reference tendency than the DNN': wasn't this already the conclusion of Section 4.1?

  We agree, but Section 4.1 focuses on the zonal wind tendency, and Section 4.3 discusses the other tendencies. We propose to modify the sentence by "As was the case with the zonal wind tendency, the U-Net architecture has a better ability to capture the reference tendencies than the DNN".

- line 465: I don't understand this sentence.

  We propose making some changes to clarify the following sentence "We refer to this laplacian as a hidden variable, often known as a latent variable, because it is obtained directly from the zonal wind", here is the proposal: "This laplacian is directly inferred from the zonal wind, hence we can refer to this variable as a hidden variable, often called latent variable".

- line 485: Did the model converge during the 'training with Δ' experiment using the same settings as in the 'initial training' experiment?

  Models do not converge perfectly because during training, we use the early stopping criterion to prevent overfitting.This information is now added in Section 3.1 "Neural network architecture" thanks to the following sentence: "It should be noted that this stopping criterion may prevent models from converging perfectly, the key point is to avoid overfitting. Thus, the models saved and used are those that achieved the best performance according to the computed loss metric on the validation dataset tracked during training".

- line 499-501: Is it a global R^2 score or only for the wind components?

  This section focuses only on the zonal wind tendency: all the scores mentioned correspond to this tendency. Following this comment, we have added "for the zonal wind tendency" to "When comparing the values of the coefficient of determination [...]" to make the text clearer.